

Extraction of Physical Effects Based on the Semantic Analysis of the Patent Texts

Marina Fomenkova^(✉), Dmitriy Korobkin^(iD),
and Sergey Fomenkov^(iD)

Volgograd State Technical University, Volgograd, Russia
mfa92@yandex.ru

Abstract. The paper represents new methodology of semantic analysis for physical effects extracting. This methodology is based on the Tuzov ontology that formally describes the Russian language. In this paper, semantic patterns were described to extract structural physical information in the form of physical effects. A new algorithm of text analysis was described. The approach is applied to the database of physical effects and to the patent texts. The results of the proposed method compared with the results of the IOFEE system that is used for the same tasks. The method described in this article allowed increasing efficiency of the physical effect elements extracting. The semantic analyzer based on the Tuzov ontology was created to increase the accuracy and completeness of the method.

Keywords: Semantic analysis · Physical effect · Fact extraction · Tuzov ontology

1 Introduction

Currently, the semantic analysis of the text is one of the most important areas in computer science. This direction is widely used in various spheres, such as trade (analysis of consumer preferences), political science and sociology (forecasting of the election results, analysis of sociological data, etc.), psychiatry (patients diagnosing), etc. However, despite its wide application, semantic text analysis is one of the most complicated mathematical problems, which consists of several stages of text processing.

In this paper, text analysis is considered as a tool for extracting structured physical knowledge in the form of physical effects (PE) [1]. Currently, the task of identifying the descriptions of PE from the texts of scientific documents for replenishing the PE base as an information basis for the development of new technical solutions is very important.

Today, there is the problem of the effectiveness of the descriptions of physical effects identifying in the texts of physical documents. Identification of descriptions of physical effects, for example, from the text of a patent is necessary to check the new incoming physical effect on the coincidence with patents existing in the database. If the descriptions of physical effects coincide in the claimed patent with descriptions in the patents from the database, the refusal in the patent application is formed.

At the moment, the system for detecting descriptions of physical effects from texts based on “Semantix” semantic analyzer has been implemented [2, 3]. The disadvantage of this system is often the incorrect construction of the semantic network, due to which individual elements of physical effects can be extracted erroneously.

It is necessary to increase the effectiveness of the descriptions of physical effects identifying. In this paper, we propose a technique for identifying descriptions of physical effects from texts of documents based on the semantic templates built on the basis of the Tuzov’s ontology.

2 Analysis of Software Systems and Tools for the Semantic Analysis of Text

An analysis of the effectiveness of software systems for extracting facts is presented in [3]. As shown in the paper, most existing systems do not have sufficient flexibility to customize on the subject area. Therefore, there is a need for a new approach to solving the problem of extracting elements of the PE description from the texts of scientific documents.

Analysis of the effectiveness of software systems for semantic analysis is shown in Tables 1 and 2. In these tables different program systems and tools for semantic analyses are presented. The effectiveness and semantic abilities are presented.

There are the efficiency criteria and the main points for analyzing and evaluating the systems for extracting knowledge and facts from the text sources. The following points were identified in the analysis of the subject area:

- Algorithm for identifying target entities (can be search by patterns or search based on neural network technologies).
- System flexibility - the ability to configure system parameters, the ability to add new entities to the system, etc. (These parameters can be set rigidly in the system, and can be set by the user of the system).
- License (closed or open).
- The accuracy of the retrieval of the facts sought - the percentage of correctly identified semantic units (1 - from 85%, 2 - 75–85%, 3 - below 75%).
- Completeness of extraction - the percentage of correctly extracted semantic units extracted from the total number of facts in the text (1- from 50%, 2 - to 50%).

As it can be seen from the Table 1 existing systems (except IOFFE) do not fit in the field of “Physics”; in the latter, the accuracy and completeness of the extraction of the elements of the PE structure are not high enough.

From the can be seen that these systems are oriented to the analysis of the any subject texts, they do not have the possibility of an advanced tuning to extract certain semantic structures, which significantly reduces the possibility of using them for solving the task. The most promising systems are the systems with the ability to configure flexibly templates for the subject area to improve the accuracy and completeness of retrieving the necessary information.

Table 1. Analysis of systems for extracting facts from text

Program system	An algorithm for target entities identifying	The system's flexibility (the ability to customize settings, adding new entities)	License	Accuracy extracting (1 - more than 75%, 2 - 65%, 3 - 65% less)	Completeness (1 - more than 60%, 2 - less than 60%)
Extracting facts from the text files, RCO [2]	Patterns search	The parameters are hard-coded	Close	1	2
Attensity text analytics [3]	Neural network technology	The parameters are hard-coded	Close	2	1
NetOwl extractor [4]	Neural network technology	The parameters are hard-coded	Close	1	1
IOFFE [1]	Patterns search	The ability to add new patterns	Open	3	2

Table 2. Semantic analyses systems

Software	Russian support	License	Base technology	Flexibility (focus on the subject area, the ability to customize, code availability)
Stanford nlp [5]	No	Open	Machine learning technologies	No
Malt parser [6]	Yes	Open	Machine learning technologies	No
Link grammar parser [7]	No	Open	Relations grammar	No
AGFL [8]	Yes	Open	AGFL-grammar	No
Tomita parser [9]	Yes	Open	CF-grammar and key words	No

3 Semantic Text Analysis to Identify Descriptions of Physical Effect

3.1 The General Approach for PE Extracting

The general approach for PE extracting is following:

- Tokenization. At this stage, sentences are broken up into words, extra characters are deleted from the sentence.
- Morphological analysis. At this stage, all the lexemes allocated during the tokenization stage are subjected to morphological analysis. Morphological analysis is performed using a TreeTagger library. This tool is a language-independent tool for morphological markup of texts. Helmut Schmitt at the Institute of Computer Linguistics at the University of Stuttgart developed it. TreeTagger has successfully proved itself in various tasks of text processing in various languages (Russian, English, German, French, Italian, etc.).

The result is the structure that consists of three columns. The first column of the TreeTagger result is the original word received at the stage of tokenization. The second column presents the morphological characteristics of the word in the sentence (gender,

number, case). In the third column, the word is presented in its initial form. If TreeTagger could not find the initial form of the word, the word <unknown> is written in the third column.

- Lemmatization. At this stage, words that were not recognized at the morphological analysis stage (the initial form of the word was not recognized) enter the input to the CstLemma application.
- Extraction of structured physical information in the form of physical effects based on semantic templates, developed based on the Tuzov's ontology.

Ontology of the Russian language [2] - a formal description of the Russian language, proposed by V.A. Tuzov (semantic roles are determined on the basis of semantic classes and morphological information). The basis of the approach is a semantic dictionary that describes more than one hundred thousand lexical units (words and phrases), and each word is described as a semantic formula consisting of basic functions.

- The language is the algebraic system $\{f_1, f_2, \dots, f_n, M\}$, where f_i are the basic functions in the language, and M is the language structure that represents the set of basic concepts m_1, \dots, m_r , and their hierarchy.
- Any sentence of a language can be represented as a superposition of basis functions f_i , through which the words of the language are also expressed, excluding the basic concepts m_j that enter in M . Thus, sentences represent single superposition of functions, in the mathematical sense are treated as functions.
- The grammar is related to the semantics of the language, which is based on a semantic dictionary that describes more than one hundred thousand words and phrases. They are divided into 3 levels: fundamental (consists of 1500 hierarchical classes, as well as a set of basic functions); Variable (consists of 23000 classes, connected with the fundamental, because they are described on the basis of this level and are its variations); Descriptive (words are described based on words and concepts of the first two levels). Each word is described as a semantic formula, which consists of basic functions.

The dictionary of basic concepts of the Russian language contains words that can not be expressed through other simpler concepts. It contains about 18,000 nouns that call physical and abstract objects, more than a thousand basic adjectives and about a thousand basic verbs, which, ultimately, have been replaced by verbal nouns. The remaining words - more than 90,000 words - are derived, that is, their meaning is expressed in the form of a superposition constructed from basic functions and basic concepts. The whole set of concepts is divided into a hierarchical class system.

EXPOSURE \$15142 (Caus_o (AGENT: SOMETHING \$ 1 ~ Gen, Lab (OBJECT:! Acc, LOCATION: Prep)))

In [9, 10] a formal description of the physical effect is presented. The physical effect is an objective, naturally conditioned connection between two or more physical phenomena, each of which is characterized by a corresponding physical quantity. From the content side, the PE is represented as a functional connection (arising from physical laws, their consequences) between two or more physical quantities.

To extract descriptions of physical effects, a model for representing structured physical information in a natural language text has been developed.

$$M_{PE} = \langle C, D, B, R_C, R_B \rangle, \quad (1)$$

where, C is the set of predicates (relations) that are characteristic for describing the PE in the text, $c_i \in C$; D - semantic roles and cases of arguments with predicates. $d_i \in D$ - list of roles/ cases of arguments, consistent with the predicate c_i ; $d_j \in D$; B - the set of elements of the description of the PE (A, B, C), $B_k \in B$.

$$\forall c_i \in C \exists d_j \in D_i d_j \longrightarrow def B_k,$$

where, $B_k \in \{input(A), output(C), object(B)\}$; def is the operator associating the role/case of the argument d_j with the predicate c_i , the set of the elements of the PE.

R_C is the relation on $C \times D$, the pair $(c_i, d_j) \in R_C$ uniquely identifies the element (s) of the PE description, consistent with the predicate role/case d_j .

R_B is the relation on $R_C \times B$, the pair $((c_i, d_j), B_k) \in R_B$ defines the set of software concepts corresponding to the element of the PE description B_k .

Let us consider an example of the application of the model.

The text containing the description of the PE: “The effect of a magnetic field on amperage in a conducting layer”.

Where $C_1 = IMPACT$

$D = \{AGENT/Generic, OBJECT/VIN, LOCATION/Prev\}$

$B = \{InputPE, OutputPE, ObjectPE\}$

For the role “Agent” of the conceptual relationship c_1 , which is performed by the member of the relation m_1 : $z_1 = \{input\ of\ the\ PE\}$.

For the role of “Object”, which is performed by the relation member m_2 : $z_2 = \{output\ of\ the\ PE, object\ PE\}$.

For the role of “Place”, which is performed by the relation member m_3 : $z_3 = \{input\ PE, object\ PE\}$. $B_1 = InputPE = Magnetic\ field$ (Exposure (of what ?, gen), AGENT),

$B_2 = OutputPE = Amperage$ (Impact (what ?, acc), OBJECT),

$B_3 = ObjectPE = conductive\ layer$ (Impact (in what ?, prep), LOCATION)

The dictionary of templates contains a set of key words (predicates), which correspond to individual sets of links “Element of the description of the PE” – “Semantic role of the argument in the Tuzov ontology”.

Predicates are the basic verbs or verbal nouns that characterize the descriptions of physical effects in the text, such as interaction, interaction, influence, impact, highlight, act, dependence, etc.

Let us consider an example of the application of the model.

The text containing the description of the PE: “The influence of the magnetic field on the Amperage in the conductive layer”, where the predicate “EFFECT”, the “Input of the PE” according to the template dictionary corresponds to the semantic role of the “Agent”, the “Output of the PE” is the Object, the “Object of the PE” is the Place. We obtain the extracted elements of the PE description:

PE Input = Magnetic field (Exposure (AGENT)), PE Output = Amperage (Impact (OBJECT)),

PE object = conductive layer (Impact (LOCATION))

Based on the semantic roles in Tuzov's ontology, the structure of the semantic template was defined:

<Predicate> <input of the PE in the required case> <preposition of the output of the PE> <output of the PE in the required case> <preposition of the PE object> <PE object in the desired case>, where predicate is the keyword, selected according to the subject domain, is a verb or verbal noun (interaction, interaction, influence, allocation, isolate, action, act, depend, dependency, etc.).

In the event that there is no preposition or semantic role for the keyword in the ontology, the word null is written in the template. In the event that several prepositions and/or cases corresponded to a single semantic role, prepositions and/or cases were recorded through separators.

On the basis of the analysis of the subject area "Physics", the keywords (predicates) most frequently encountered in conjunction with unstructured descriptions of physical effects were identified. Such keywords are either verbs or other verbal parts of speech (noun, participle).

As a result, more than 100 keywords were chosen for the subject area.

Based on the semantic roles in Tuzov's ontology, the structure of the semantic template was defined:

<Predicate> <input of the PE in the required case> <preposition of the output of the PE> <output of the PE in the required case> <preposition of the PE object> <PE object in the desired case>, where predicate - the keyword, selected according to the subject domain, is a verb or verbal noun (interaction, interaction, influence, influence, influence, influence, allocation, isolate, action, act, depend, dependency, etc.);

In the event that there is no preposition or semantic role for the keyword in the ontology, the word null is written in the template. In the event that several prepositions and/or cases corresponded to a single semantic role, prepositions and/or cases were recorded through separators.

As a result, more than 100 semantic templates for the subject area were developed.

Algorithm of semantic analysis for extracting arguments Agent, Object, Place is represented in Fig. 1.

Thus, the algorithm consists of the following basic steps:

- Tokenization;
- Morphological analysis;
- Lemmatization;
- Extract the arguments of the semantic roles of the Agent, Object and Place - nouns and adjectives related to the given nouns. In addition, connections were extracted from cases - a connection was made in the genitive case, consistent with the noun corresponding to the semantic role.

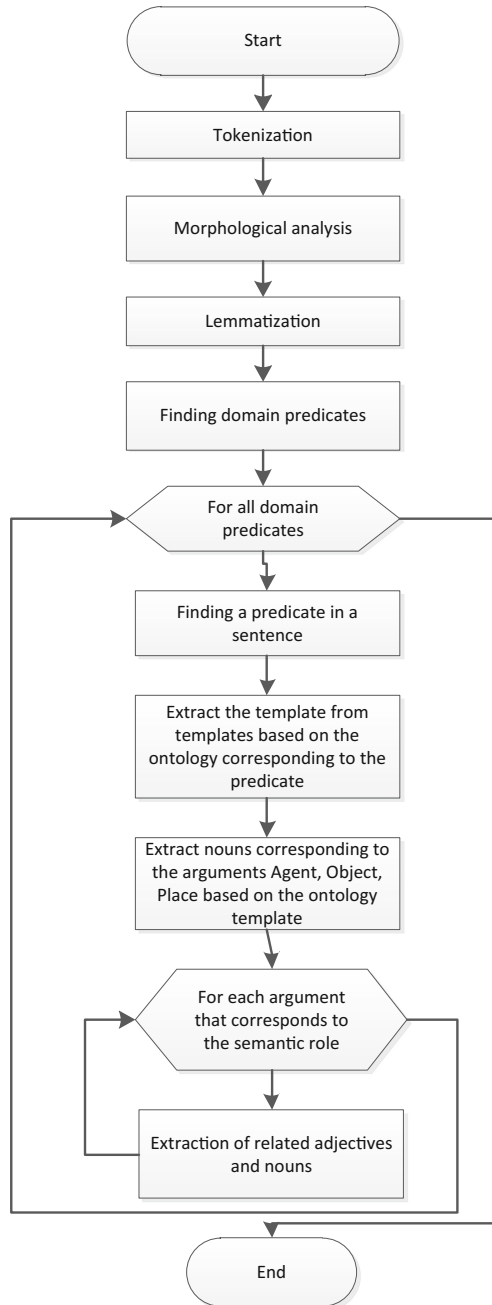


Fig. 1. Algorithm of semantic analysis for extracting arguments Agent, Object, Place

3.2 Extraction of Structured Physical Information in the Form of Physical Effects and Removal of Semantic Ambiguity. Algorithm for Constructing a Match Template

Because of the analysis of the subject domain, a phenomenon of semantic ambiguity arises. Semantic ambiguity arises when it is possible to match one semantic role in ontology. Tuzov (“Agent”, “Object”, “Place”) to several elements of the description of the physical effect (“Input”, “Exit”, “Object” of the physical effect).

For example, if we consider the following excerpt of a sentence containing a physical effect:

“IMPACT ON THE SEMICONDUCTOR”

The word “semiconductor” in this case is the object of physical effect. The semantic role in Tuzov’s ontology for this word is similar – “Object”.

If we consider another passage of a sentence containing a physical effect:

“IMPACT ON THE MAGNETIC FIELD”

The word “magnetic field” in this case is the output of the physical effect. The semantic role in the ontology of Tuzov is still the “Object”.

Thus, one semantic role in the ontology of Tuzov can be compared, depending on the context, several elements of the description of physical effects. This affects the correctness of determining the elements of physical effects according to the developed semantic templates.

To remove this kind of semantic ambiguity, a separate program module was developed, which is necessary to compare with the existing descriptions of physical effects and semantic roles described in templates based on the ontology of Tuzov. The essence of the approach is to compare the semantic roles of Tuzov, elements of the description of physical effects and specific words with which ambiguity can be revealed. Based on the analysis of the “Description of the PE” field of the description of 1200 physical effects, there is a correspondence between the semantic role and the element of the PE description. The template for extracting the PE is formed based on the statistics of this bundle for each predicate.

To eliminate the ambiguity based on the field “Description of the PE” of the database table and Tuzov ontology (predicates, cases and prepositions for semantic roles) is the correspondence:

<Semantic role> - <Element of the description of PE>.

Based on the statistics of this bundle, a template for extracting the PE is formed for each predicate.

Thus, a set of words that correspond to different elements of the description of physical effects is defined. For example, for the above example with the word “Impact”, this keyword is the word “Semiconductor”, and the output of the physical effect is the word “Magnetic field”.

Algorithm for constructing correlations between the semantic roles of the Tuzov ontology and elements of the description of the physical effect for predicates of the domain are shown in Fig. 2.

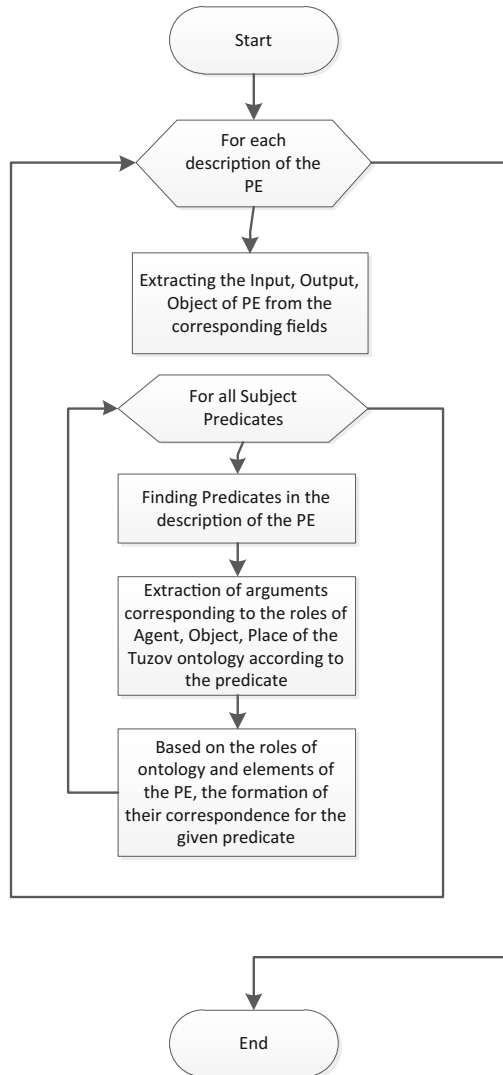


Fig. 2. The algorithm of construction of correspondences to semantic roles of Tuzov ontology to elements of the description of physical effect for predicates of the subject domain

So the essence of the approach is to compare the semantic roles of Tuzov, the elements of describing physical effects and specific words with which ambiguity can appear. To eliminate the ambiguity based on the field “Description of the PE” of the database table and Tuzov ontology (predicates, cases and prepositions for semantic roles) is the correspondence:

<Semantic role> - <Element of the description of PE>.

Based on the statistics of this bundle, a template for extracting the PE is formed for each predicate.

Thus, a set of words that correspond to different elements of the description of physical effects is defined.

The general algorithm for extracting physical effects is shown in Fig. 3.

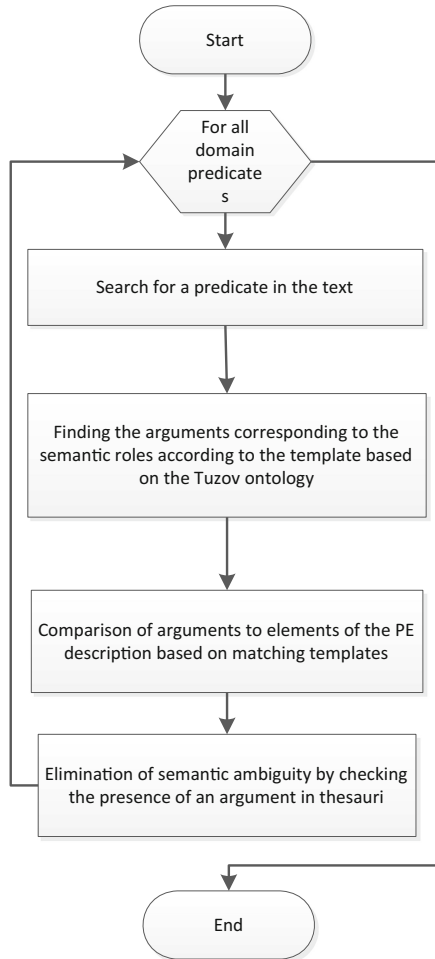


Fig. 3. The algorithm for extracting elements of physical effects

Thus, the algorithm for extracting elements of physical effects reduces to finding all the predicates of the domain and performing the following steps for them:

- Search for a predicate in the text;
- Finding the arguments corresponding to the semantic roles according to the template based on Tuzov's ontology;

- Comparison of arguments to elements of the description of the physical effect on the basis of match patterns;
- Elimination of semantic ambiguity by checking the presence of an argument in thesauri.

4 Results

The following performance criteria were developed:

- Accuracy of extraction of elements of PE description;
- Completeness of extraction of elements of PE description;
- F-measure.

The accuracy is characterized by the number of correctly extracted elements to the total number of elements of the PE description

$$P = N_r / N_f, \quad (2)$$

where, P is the accuracy of the PE extraction, N_r is the number of correctly extracted elements of the PE description, N_f is the number of elements of the FE description found in the text.

The completeness of the PE description elements extraction is a value expressed in percent, which characterizes the number of elements found in the PE description to the total number of elements of the PE description in the text.

$$R = N_f / N, \quad (3)$$

where, R is the completeness of the extraction of elements of the PE description, N_f is the number of elements of the PE found in the text, N is the total number of elements of the PE description in the text.

The F-measure is calculated by formulas 4 and 5.

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}, \quad (4)$$

$$\beta^2 = \frac{1 - a}{a}, \quad (5)$$

The data was tested based on physical effects data, developed at the CAD department of Volgograd State Technical University. 100 physical effects were selected. The tests were carried out on the descriptions of physical effects in the DB of the PE and the results of the extraction were compared with the fields “Input”, “Output”, “Object” of the physical effect.

Also, the tests were conducted on 31 patent documents for which physical effects contained in the patent description are known in advance.

The results were compared with the results of the IOFFE program [11] based on “Semantix” semantic analyzer.

The efficiency analysis showed that the developed system increases the efficiency of extracting elements of the description of physical effects by 4% for accuracy and by 7% for completeness (Tables 3 and 4).

Table 3. Analysis of efficiency using the PE database

	Accuracy (%)	Completeness (%)	F-measure
Our system	68	59	61.5
IOFFE	57	53	54.14

Table 4. Analysis of the effectiveness using the patent array

	Accuracy (%)	Completeness (%)	F-measure
Our system	53	46	47.9
IOFFE	46	41	42.38

Sample 1. Patent description: “The photoelectric conversion element may be a photodiode having a p-n junction or a pin junction, a phototransistor, or the like. When the incident light hits the semiconductor junction of the cell, this light leads to the appearance of the photoelectric effect, in which electric charges arise.” [14].

- PE Input: “Light, any other electromagnetic radiation (energy - eV)”. PE Output: “The electric charge (electron emission), (J)”.
- PE Object: “Photoconductive material (photoconductor)”.

The results of the program show the results of the physical effect elements extraction:

- PE Input: “Light”.
- PE Output: “electric charge”. PE Object: “phototransistor”.

Sample 2. Patent description: “In electrical circuits, any electric current produces a magnetic field and hence generates a total magnetic flux acting on the circuit”.

- PE Input: “Electric current”.
- PE Object: “Electrical circuit”.
- PE Output: “Magnetic field”.

The results of the program show the results of the physical effect elements extraction:

- PE Input: “Electric current”.
- PE Object: “Electrical circuit”.
- PE Output: “Magnetic field”.

5 Conclusion

The method described in this article allowed increasing efficiency of the PE elements extracting. The semantic analyzer based on the Tuzov ontology was created to increase the accuracy and completeness of the method. The approach was tested on the PE database and the patent array.

Key words - predicates - (verbs, verbal nouns, participles) that can be found in sentences containing elements for describing physical effects have been identified. The corresponding descriptions are selected in the ontology.

On the basis of the ontology of Tuzov, semantic templates have been developed that consist of a keyword (predicate) - a characteristic verb or verbal form for a given field, cases of the semantic roles “Agent”, “Object”, “Place”, and related prepositions. Then the templates of the correspondence of the semantic roles of the ontology of Tuzov and the elements of the description of physical effects are constructed. Based on the thesauri of the subject area, this allows you to extract elements of the description of physical effects from text documents.

The efficiency analysis technique included finding the accuracy and completeness of extracting elements of descriptions of physical effects from the field of describing physical effects in the database, as well as in the texts of the patent documents. Then the results were compared with the results of the IOFFE system, which was implemented based on Semantix semantic analyzer. Thus, the efficiency of the developed software was evaluated.

The results of the efficiency analysis showed that the accuracy of extraction of elements of the description of physical effects was increased by 11% for physical effects from the database and 5% for physical effects from the texts of patent documents. Completeness of extraction of elements of the description of physical effects was increased by 6% for physical effects from the database and by 7% for physical effects from the texts of patent documents.

Acknowledgement. This research was partially supported by the Russian Fund of Basic Research (grants No. 15-07-09142 A, No.15-07-06254 A, No. 16-07-00534 A).

References

1. Korobkin, D.M., Fomenkov, S.A., Kamaev, V.A., Fomenkova, M.A.: Multi-agent model of ontology- based extraction of physical effects descriptions from natural language text. In: Information Technologies in Science, Management, Social Sphere and Medicine (ITSMSSM), pp. 498–501. Atlantis Press (2016). doi:[10.2991/itsmssm-16.2016.13](https://doi.org/10.2991/itsmssm-16.2016.13)
2. Pleshko, V.V., Ermakov, A.V.: Semantic interpretation in text analysis computer systems. J. Inf. Technol. (2009)
3. Taylor, J.: First Look – Attensity 5.5 decision, management, solutions. In: 5 International Conference on Information Technologies in Business and Industry (2017)
4. Krupka, G.R., Hausman, K.I.: Description of the NetOwl TM Extractor system as used for MUC-7. In: MUC-7 (1998)

5. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press, Cambridge (2008)
6. Nivre, J., Hall, J., Nilsson, J.: MaltParser: a data-driven parser-generator for dependency parsing. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy, pp. 2216–2219 (2006)
7. Tukeyev, U.A., Melby, A.K., Zhumanov, Z.M.: Models and algorithms of translation of the Kazakh language sentences into English language with use of link grammar and the statistical approach. In: *IV Congress of the Turkic World Math. Society*, Baku (2014)
8. Azarova, I.: The matching of AGFL subcategories to Russian lexical and grammatical groupings. In: *Proceedings of the Second AGFL Workshop on Syntactic Description and Processing of Natural Language* (2002)
9. Ogogrodnik, P.B., Serebryannaya, L.V.: Text analysis with Tomita parser. In: *International Conference (BSUIR)*, Minsk, pp. 230–231 (2014)
10. Tuzo, V.A.: *Computer Linguistics*. St. Petersburg State University Press, St. Petersburg (1998)
11. Korobkin, D., Fomenkov, S., Kolesnikov, S., Lobeyko, V., Golovanchikov, A.: Modification of physical effect model for the synthesis of the physical operation principles of technical system. In: Kravets, A., Shcherbakov, M., Kultsova, M., Shabalina, O. (eds.) *CIT&DS 2015*. CCIS, vol. 535, pp. 368–378. Springer, Cham (2015). doi:[10.1007/978-3-319-23766-4_29](https://doi.org/10.1007/978-3-319-23766-4_29)
12. Fomenkov, S., Korobkin, D., Kolesnikov, S.: Method of ontology-based extraction of physical effect description from Russian text. In: Kravets, A., Shcherbakov, M., Kultsova, M., Iijima, T. (eds.) *JCKBSE 2014*. CCIS, vol. 466, pp. 321–330. Springer, Cham (2014). doi:[10.1007/978-3-319-11854-3_27](https://doi.org/10.1007/978-3-319-11854-3_27)
13. Fomenkov, S.A., Kolesnikov, S.G., Korobkin, D.M., Kamaev, V.A., Orlova, Y.A.: The information filling of the database by physical effects. *J. Eng. Appl. Sci.* **9**, 422–426 (2014)
14. Patent US 20100051095 A1 “Hybrid Photovoltaic Cell Using Amorphous Silicon Germanium Absorbers With Wide Bandgap Dopant Layers and an Up-Converter”
15. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *J. Inf. Process. Manag.* **24**(5), 513–523 (1988)
16. U.S. Patent Grant Data/XML v4.3. http://www.uspto.gov/sites/default/files/products/Patent_Grant_XML_v4.3.pdf
17. Zlotin, B., Zusman, A.: *Directed Evolution: Philosophy Theory and Practice*. Ideation International, Southfield (2001)
18. Fey, V., Rivin, E.: *Innovation on Demand: New Product Development Using TRIZ*. Cambridge University Press, Cambridge (2005)
19. Blei, D.M.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**(4–5), 993–1022 (2003)
20. Korobkin, D.M., Fomenkov, S.A., Kravets, A.G., Golovanchikov, A.B.: Patent data analysis system for information extraction tasks. In: *13th International Conference on Applied Computing (AC)*, pp. 215–219. IADIS (2016)
21. Toutanova, K., Manning, C.D.: Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: *EMNLP 2000*, Hong Kong, pp. 63–70 (2000)
22. Hall, J.: *MaltParser – An Architecture for Inductive Labeled Dependency Parsing*. University of Colorado (2006)
23. Mel’čuk, I.A.: *Dependency Syntax: Theory and Practice*. SUNY Publication, New York (1988)
24. Stanford typed dependencies manual. https://nlp.stanford.edu/software/dependencies_manual.pdf

25. Korobkin, D.M., Fomenkov, S.A., Kolesnikov, S.G., Voronin, Y.F.: System of physical effects extraction from natural language text in the internet. *J. World Appl. Sci. J.* **24**, 55–61 (2013)
26. Dvoryankin, A.M., Polovinkin, A.I., Sobolev, A.N.: Automating the search for operation principles of technical systems on the basis of a bank of physical phenomena. *J. Cybern.* **14**, 79–86 (1978)
27. Arel, E.: *Goldfire Innovator. Volume II: Patents and Innovation Trend Analysis User Guide*. Invention Machine Corporation, Boston (2004)